



How Does the Degree of Valence Influence Affective Auditory P300-Based BCIs?

Akinari Onishi^{1*} and Seiji Nakagawa^{1,2,3}

¹ Center for Frontier Medical Engineering, Chiba University, Chiba, Japan, ² Department of Medical Engineering, Graduate School of Engineering, Chiba University, Chiba, Japan, ³ University Hospital Med-Tech Link Center, Chiba University, Chiba, Japan

OPEN ACCESS

Edited by:

Gerwin Schalk,
Wadsworth Center, United States

Reviewed by:

Fabien Lotte,
Institut National de Recherche en
Informatique et en Automatique
(INRIA), France
Dan Zhang,
Tsinghua University, China

*Correspondence:

Akinari Onishi
a-onishi@chiba-u.jp

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 01 August 2018

Accepted: 17 January 2019

Published: 19 February 2019

Citation:

Onishi A and Nakagawa S (2019) How
Does the Degree of Valence Influence
Affective Auditory P300-Based BCIs?
Front. Neurosci. 13:45.
doi: 10.3389/fnins.2019.00045

A brain-computer interface (BCI) translates brain signals into commands for the control of devices and for communication. BCIs enable persons with disabilities to communicate externally. Positive and negative affective sounds have been introduced to P300-based BCIs; however, how the degree of valence (e.g., very positive or positive) influences the BCI has not been investigated. To further examine the influence of affective sounds in P300-based BCIs, we applied sounds with five degrees of valence to the P300-based BCI. The sound valence ranged from very negative to very positive, as determined by Scheffe's method. The effect of sound valence on the BCI was evaluated by waveform analyses, followed by the evaluation of offline stimulus-wise classification accuracy. As a result, the late component of P300 showed significantly higher point-biserial correlation coefficients in response to very positive and very negative sounds than in response to the other sounds. The offline stimulus-wise classification accuracy was estimated from a region-of-interest. The analysis showed that the very negative sound achieved the highest accuracy and the very positive sound achieved the second highest accuracy, suggesting that the very positive sound and the very negative sound may be required to improve the accuracy.

Keywords: BCI, BMI, P300, EEG, affective stimulus

1. INTRODUCTION

A brain-computer interface (BCI), also referred to as a brain-machine interface (BMI), translates human brain signals into commands that can be used to control assistive devices or communicate externally with others (Wolpaw et al., 2000, 2002). For instance, BCIs are used to help persons with disabilities interact with the external world (Birbaumer and Cohen, 2007). BCIs decode brain signals obtained by invasive measurements such as electrocorticography or by non-invasive measurements such as scalp electroencephalography (EEG) (Pfurtscheller et al., 2008). One of the most successful BCIs utilizes EEG signals in response to stimuli – i.e., event-related potentials (ERPs). In particular, when a subject discriminates a rarely encountered stimulus from a frequently encountered stimulus, the ERP elicits a positive peak at around 300 ms from stimulus onset; this peak is named the P300 component. ERPs with P300 have been applied to BCIs, and such a system is called a P300-based BCI (Farwell and Donchin, 1988).

Early studies of the P300-based BCI commonly used visual stimuli. For instance, Farwell and Donchin (1988) proposed a visual P300-based BCI speller that intensified letters in a 6 × 6 matrix

by row or by column (row-column paradigm). Townsend et al. (2010) proposed an improved P300 speller that highlighted letters in randomized flashing patterns derived from a checkerboard, which was called the checkerboard paradigm. Ikegami et al. (2014) developed a region-based two-step P300 speller that could use the Japanese Hiragana syllable, and took advantage of a green/blue flicker and region selection. In addition, facial images (e.g., famous faces and smiling faces) were applied to visual P300-based BCIs, which resulted in innovative improvement (Kaufmann et al., 2011; Jin et al., 2012, 2014). Although visual P300-based BCIs have been improved, and have shown to be effective in clinical studies, they still depend on the eye condition and can be affected by whether the eyes are open or closed, limited eye gaze, or limited sight. Therefore, it is important that other sensory modalities be studied and improved upon.

In addition to visual stimuli, other modalities, for example auditory stimuli, have been studied. In an early study of auditory P300-based BCIs, auditory stimuli, such as the sound of the word “yes,” “no,” “pass,” and “end” were assessed in a BCI developed by Sellers and Donchin (2006). Klobassa et al. (2009) examined the use of bell, bass, ring, thud, chord, and buzz sounds. Moreover, by pairing numbers with letters using a visual support matrix, Furdea et al. (2009) succeeded in spelling letters by counting audibly pronounced numbers. In addition, beep sounds were used as auditory stimuli and the effects of pitch, duration, and sound source direction were assessed by Halder et al. (2010). Höhne et al. (2010) developed a two-dimensional BCI that varied in sound pitch and location of sound source. Although the auditory BCI is advantageous because it is independent of eye gaze, auditory BCIs have shown worse performance than visual BCIs in several studies (Sellers and Donchin, 2006; Wang et al., 2015), and it is clear that improvements to the auditory BCI are required. Thus, we focused on improving auditory stimuli for BCIs in the current study.

The auditory P300-based BCI can be improved by using sophisticated stimuli. Schreuder et al. (2010) demonstrated that presenting sounds from five speakers in different locations resulted in better BCI performance than that obtained using a single speaker. Additionally, Höhne et al. (2012) evaluated spoken or sung syllables as natural auditory stimuli, and compared to the artificial tones, found that the use of natural stimuli improved the users' ergonomic ratings and the classification performance of the BCI. Simon et al. (2014) also applied natural stimuli of animal sounds. Guo et al. (2010) employed sound involving spatial and gender properties together with discriminating properties of sounds (active mental task), which served to enhance the late positive component (LPC) and N2. Recently, Huang et al. (2018) explored the use of dripping sounds and found that the BCI classification accuracy was higher than when beeping sounds were used. We previously applied sounds with two degrees of valence (positive and negative) to the auditory P300-based BCI (Onishi et al., 2017). We confirmed the enhancement of the late component of P300 in response to those stimuli. However, how degrees of valence (e.g., very positive or positive) influence the P300-based BCI remains unknown.

This study aimed to clarify how degrees of valence in sounds influence the auditory P300-based BCIs. Five sounds with

different degrees of valence were applied to the P300-based BCI: very negative, negative, neutral, positive, and very positive. We hypothesized that the valence should exceed a certain degree because the amplitude of P300 was not in proportion to the emotion (Steinbeis et al., 2006). Since the auditory P300 BCI requires a larger amount of training data, and can cause fatigue when applied to the BCI separately, we applied these sounds to the BCI together. The influence caused by the sound valences was then analyzed offline with cross-validation. To confirm the valence of those sounds, the Scheffé's method of paired comparison was applied. We also performed a waveform analysis using the point-biserial correlation coefficient to reveal ERP components that contributed to the classification. Based on the waveform analysis, a region-of-interest (ROI) was identified, and then specially designed cross-validation was performed to estimate offline stimulus-wise classification accuracy. The online performance of the BCI and its preliminary feature extraction process has previously been demonstrated; however, the effect of these affective sounds was not evaluated (Onishi and Nakagawa, 2018). Therefore, in the current study, the influence of affective auditory stimuli on the BCI was evaluated based on the waveform analysis and the stimulus-wise classification accuracy. In contrast to our previous study (Onishi et al., 2017), which revealed whether affective sound is effective to the BCI, the current study aimed to clarify how degrees of valence influence the BCI. The name of the late component seen in our previous study (Onishi et al., 2017) is not consistent within the literature; P300, P3b, P600, late positive component, or late positive complex are used (Finnigan et al., 2002). In the current study, we used the term late component of P300 in reference to those ERP components.

2. METHODS

2.1. Subjects

Eighteen healthy subjects aged 20.6 ± 0.8 (9 females and 9 males) participated in this study. All participants were right-handed as assessed by the Japanese version of the Edinburgh Handedness Inventory (Oldfield, 1971). All subjects gave written informed consent before the experiment. This experiment was approved by the Internal Ethics Committee at Chiba University and conducted in accordance with the approved guidelines.

2.2. Stimuli

Five cat sounds, representing five different valences (very negative, negative, neutral, positive, and very positive), were prepared. Sounds were cut to 500 ms, then rises and falls of the sounds were linearly faded in and out. The root-mean-square of each sound was equalized. Detailed conditions of sound processing were presented in **Table 1**. The sound was emitted through ATH-M20x headphones (Audio-Technica Co., Japan) via audio interface UCA222 (Behringer GmbH, Germany).

Degrees of valence for these sounds were rated and verified using Ura's variation of the Scheffé's method (Scheffé, 1952; Nagasawa, 2002). We used this method instead of visual analog scale because it provides more reliable ratings using the paired comparison of sounds. Specifically, a computer first randomly selected two sounds (sounds A and B) out of the five sounds.

TABLE 1 | Properties of affective sounds.

No.	Valence label	Sound name	Source	Trimming sample
1	Very negative	cat-fight	http://taira-komori.jp.org/	64,910
2	Negative	Cat_Meowing_2-Mr_Smith-780889994	http://soundbible.com/	12,000
3	Neutral	cat6	http://pocket-se.info/	2,769
4	Positive	catvoice	http://taira-komori.jp.org/	882
5	Very positive	kitty	http://pocket-se.info/	112

Second, a subject listened to sound A, and then sound B, only once. After listening to sounds A and B, the subject rated which sounded more positive by reporting ± 3 (+3: B is very positive, 0: neutral, -3: very negative). Participants answered the degrees of valence for a total of 20 sound pairs. The degrees of valence were statistically tested using the analysis of variance (ANOVA). The ANOVA was modeled for Ura's variation of the Scheffé's method, which contains factors of the average of ratings, the individual difference of the ratings, the combination effect, the average of the order effect, and the individual difference of the order effect. Note that the main objective is to reveal the main effect of the averaged ratings, and the others are optional factors. See more detail in (Nagasawa, 2002). The analysis was followed by a comparison of the differences of ratings between each pair using a 99% confidence interval.

2.3. EEG Recording

EEG signals were recorded using MEB-2312 (Nihon-Kohden, Japan). EEG electrodes were placed at C3, Cz, C4, P3, Pz, P4, O1, and O2 (Onishi et al., 2017) according to the international 10–20 system. The ground electrode was at the forehead and the reference electrodes were at the mastoids. A hardware bandpass filter (0.1–50 Hz) and a notch filter (50 Hz) were applied by the EEG system. The EEG was digitized using a USB-6341 (National Instruments Inc., USA). The sampling rate of the EEG analysis was 256 Hz. Data acquisition, stimulation, and signal processing were completed using MATLAB (Mathworks Inc., USA).

The experiment consisted of a training part and a testing part (see **Figure 1**). Each part contained five sessions. In each session, participants completed five runs. At the beginning of the run, participants were instructed to silently count the number of occurrences of a particular target sound, which was emitted in sequence with other non-target sounds (see **Figure 2**). Five sounds were then provided in pseudo-random order and each stimulus was repeated 10 times. The stimulus onset asynchrony was 500 ms. EEG signals were recorded during the task. In the testing sessions, outputs were estimated by analyzing EEG signals, and the output was fed back to the subject. For calculating the feedback, smoothing (4 sample window), Savitzky-Golay filter (5th order, 81 sample window), and downsampling (64 Hz) were applied before the classification by the stepwise linear discriminant analysis (SWLDA) (Krusienski et al., 2006). The online classification accuracy was 84.1% (Onishi and Nakagawa, 2018). Between sessions, subjects were asked to take a rest for a

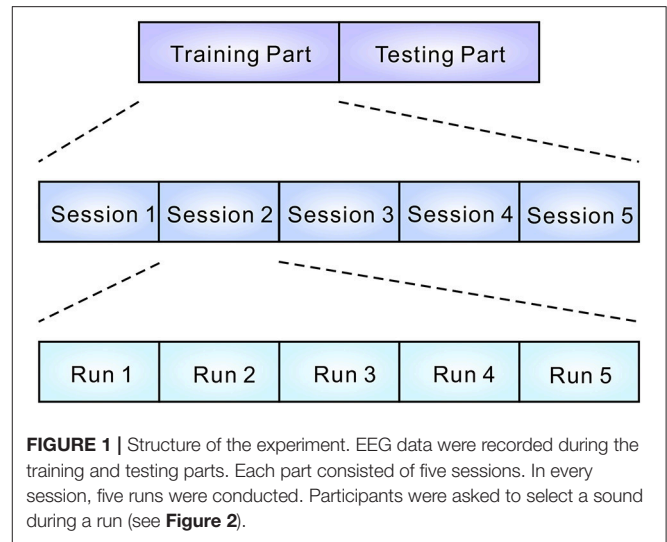


FIGURE 1 | Structure of the experiment. EEG data were recorded during the training and testing parts. Each part consisted of five sessions. In every session, five runs were conducted. Participants were asked to select a sound during a run (see **Figure 2**).

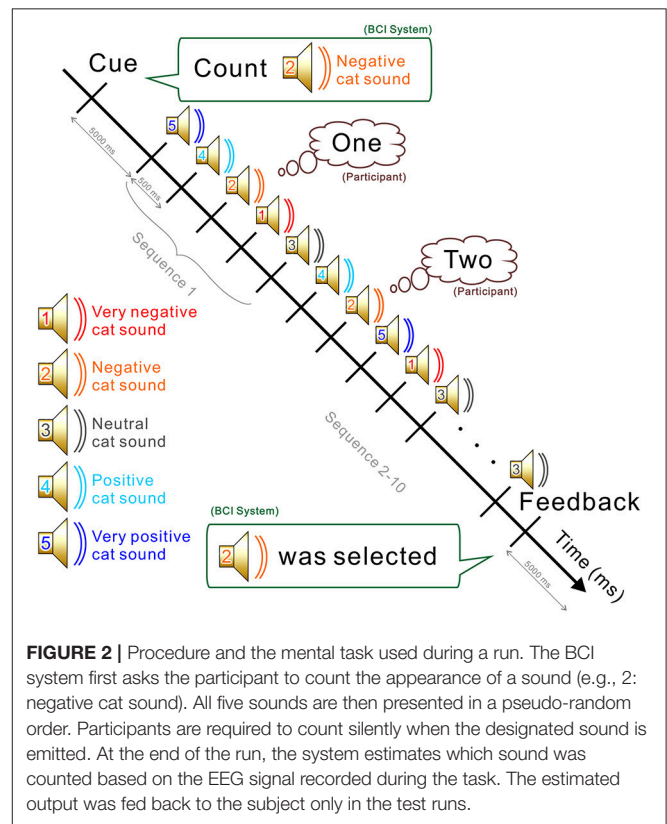


FIGURE 2 | Procedure and the mental task used during a run. The BCI system first asks the participant to count the appearance of a sound (e.g., 2: negative cat sound). All five sounds are then presented in a pseudo-random order. Participants are required to count silently when the designated sound is emitted. At the end of the run, the system estimates which sound was counted based on the EEG signal recorded during the task. The estimated output was fed back to the subject only in the test runs.

few minutes (depending on tiredness). Each sound was selected as a target once in every session. To avoid the effect of tiredness, we applied five sounds together to the BCI, then effects of valences were analyzed offline.

2.4. Waveform Analysis

Averaged ERP waveforms for each sound were estimated from the ERP data recorded during the training and testing runs

during which a sound was set as the target. The non-target averaged waveforms for each sound were also estimated. All data were preprocessed by applying the baseline removal estimated from -100 to 0 ms waveforms; a software bandpass filter (Butterworth, 6th order, 0.1 – 20 Hz) was also applied. ERPs that exceeded $80\mu\text{V}$ were removed.

In order to reveal which ERP components contributed to the EEG classification, the point-biserial correlation coefficients (r^2 values) were estimated (Tate, 1954; Blankertz et al., 2011; Onishi et al., 2017). The point-biserial correlation coefficient is defined as follows:

$$r := \frac{\sqrt{N_2 \cdot N_1}}{N_2 + N_1} \frac{\mu_2 - \mu_1}{\sigma}, \quad (1)$$

where N_2 and N_1 indicate the number of data in target (2) and non-target (1) classes, μ_2 and μ_1 are the mean value of target and non-target, and the σ denotes the standard deviation of a sample in a channel. The point-biserial correlation coefficient is equivalent to Pearson correlation between the amplitude of the ERP (continuous measured variable) and classes (dichotomous variable). The squared r (r^2) was used for the waveform analysis. It becomes higher as the difference of mean values between classes is larger and the standard deviation is smaller. We used the method instead of traditional ERP component statistics because it provides rich spatio-temporal information. The r^2 values were evaluated using a test of no correlation and p values were corrected using Bonferroni's method. If the r^2 value was not significant, the value was presented as a zero.

2.5. Classification

The offline stimulus-wise classification accuracy (Onishi et al., 2017) was computed using stimulus-wise leave-one-out cross-validation (LOOCV). First, training and testing runs, in which a sound was designated as a target, were selected from all 50 runs. Therefore, ten runs were selected in total by the procedure. Second, one run was selected as a testing run and the others as training runs. Third, a supervised classifier was trained on the ERP data during the training runs and then the test data were classified as correct or incorrect. The above two processes were repeated for all 10 runs. The classification accuracy for a sound was calculated as the percentage of correct answers during the 10 runs. We employed this LOOCV to evaluate the influence of each stimulus from the limited amount of data.

We applied baseline removal estimated from -100 to 0 ms waveforms and a software bandpass filter (Butterworth, 6th order, 0.1 – 20 Hz). They were not downsampled in order to compare the results of waveform analysis and classification. Then they were vectorized before applying classification. We used the SWLDA classifier (Krusienski et al., 2006). In summary, given the weight vector of SWLDA \mathbf{w} and the preprocessed EEG data of i -th stimulus number in s -th stimulus sequence (repetition) $\mathbf{x}_{s,i}$, the output \hat{i} can be estimated as

$$\hat{i} = \arg \max_{i \in I} \sum_{s=1}^S \mathbf{w} \cdot \mathbf{x}_{s,i}, \quad (2)$$

where $S \in \{1, 2, \dots, 10\}$ denotes the maximum number of stimuli used during the offline analysis, and $I \in \{1, 2, \dots, 5\}$ indicates the list of stimulus numbers (see **Table 1**). The offline stimulus-wise classification accuracy was calculated $\#correct\ runs / \#total\ runs$ fixing S . The thresholds of the stepwise method were $p_{in} = 0.1$ and $p_{out} = 0.15$. Training data that exceeded $80\mu\text{V}$ were removed. Testing data that showed over $80\mu\text{V}$ amplitude were set to zero to reduce the influence of outliers. The stimulus-wise classification accuracy was statistically tested by the two-way repeated-measures ANOVA, where the factors are the type of stimulus and the number of stimulus sequences (repetitions).

To identify how the components seen in the waveform analysis contributed to the classification, we calculated the stimulus-wise classification accuracy by applying a ROI. The ROI analysis is especially used in fMRI studies to clarify which area of the brain was activated (Brett et al., 2002; Poldrack, 2007). Since ERP studies focus on ERP components spread over channels, spatio-temporal ROI selection was applied in this study. Specifically, the ROI in this study was set to C3, Cz, and C4 in the 400 – 700 ms. The effect caused by the ROI was the same among comparison conditions. The similar analysis has been applied in P300-based BCI studies to reveal the effect of ERP components (Guo et al., 2010; Brunner et al., 2011). We decided to apply the ROI to reveal the effect of sound valence in response to results of waveform analysis because the automatic feature selection does not ensure which component to select and cannot support the conclusion.

3. RESULTS

3.1. Subjective Reports

Since the valences of the selected sounds were not confirmed, their valences were verified using Ura's variation of Scheffé's method. **Figure 3** shows the valence of each sound. Valences of stimulus 1 to 5 in **Table 1** were rated -1.48 , -0.72 , 0.20 , 0.75 , and 1.24 , respectively. The rated valence was analyzed by the ANOVA. It revealed the significant main effects of the average of ratings, the individual difference of the ratings, and the combination effects ($p < 0.01$). No significant main effect was seen in factors of the average of order effect ($p = 0.074$) and the individual difference of the order effect ($p = 0.509$). **Table 2** shows the 99% confidence intervals of valence estimated by Scheffé method. Since any pairs do not contain zero between upper and lower limits, all valence scores were significantly different from each other ($p < 0.01$, see **Table 1**). The result implies that the valences of the sounds were labeled properly, and they were distributed so that they are different from each other.

3.2. Waveform Analysis

The valence of the affective sounds modulated the late component of the P300 amplitude with respect to valence. **Figure 4** shows the averaged target and nontarget ERP waveforms obtained at Cz for each sound (waveforms for all channels were also presented in **Figures S1–S5**). The peak amplitude of the component was lowest in the neutral auditory stimulus (stimulus 3), while the peak was greatest for very negative and very positive auditory stimuli (stimulus 1 and 5,

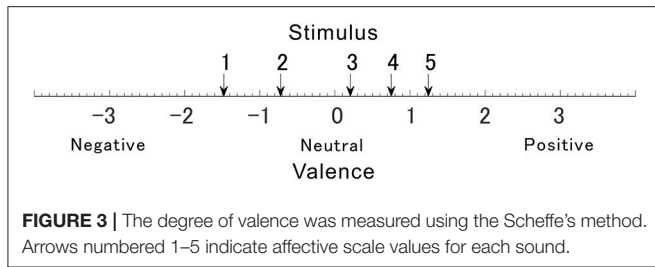


FIGURE 3 | The degree of valence was measured using the Scheffe's method. Arrows numbered 1–5 indicate affective scale values for each sound.

TABLE 2 | The 99% confidence intervals of valence estimated by Scheffe method.

Stimulus pair	Lower limit	Upper limit
1, 2	-1.00	-0.76
1, 3	-1.93	-1.68
1, 4	-2.47	-2.23
1, 5	-2.97	-2.72
2, 3	-1.17	-0.93
2, 4	-1.72	-1.47
2, 5	-2.21	-1.97
3, 4	-0.79	-0.54
3, 5	-1.28	-1.04
4, 5	-0.74	-0.49

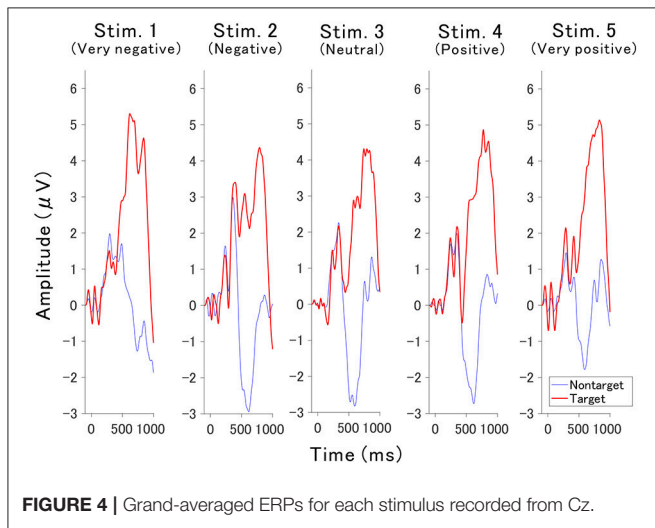


FIGURE 4 | Grand-averaged ERPs for each stimulus recorded from Cz.

respectively). To know the contribution to the classification and the statistical analysis, the point-biserial correlation coefficient analysis was applied.

The point-biserial correlation coefficients (r^2 values) provided in **Figure 5** indicate how each auditory stimulus in a channel contributes to the classification. A test of no correlation was applied to each value, then the r^2 values were set to zero if the point-biserial correlation was insignificant. The r^2 values increase around channels C3, Cz, and C4 at approximately 400–700 ms, which corresponds to the late component of P300 (**Figure 4**).

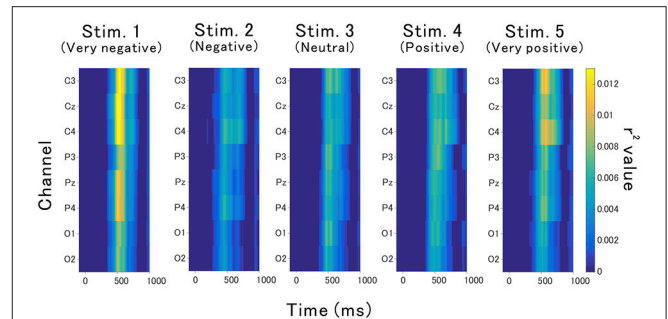


FIGURE 5 | Contribution of each feature to the classification accuracy. This contribution was estimated based on r^2 values.

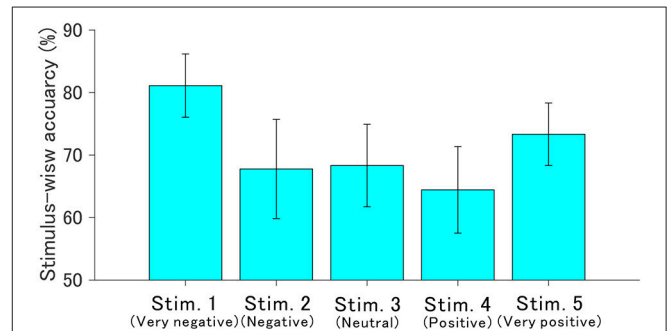


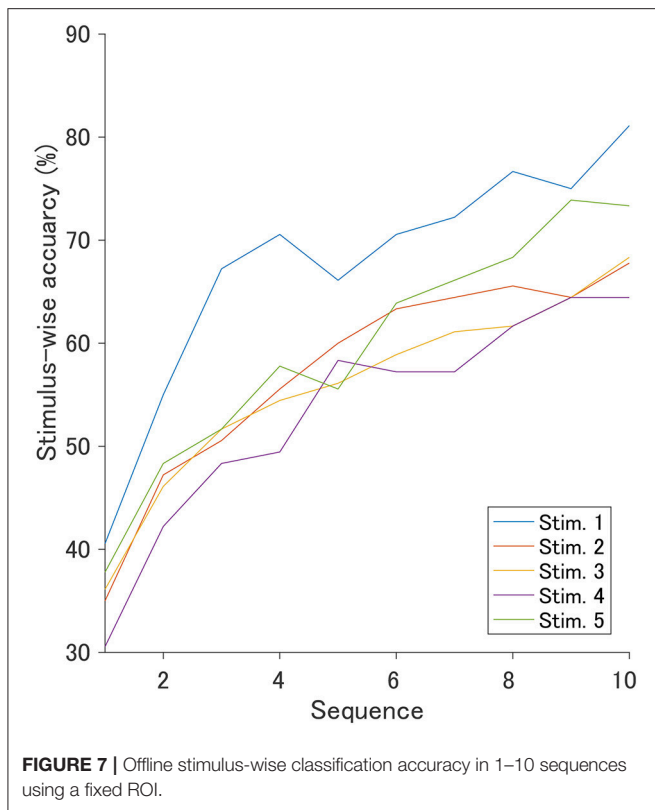
FIGURE 6 | Offline stimulus-wise classification accuracy and standard error for each auditory stimulus. These were calculated based on ERP data in the ROI (400–700 ms in C3, Cz, and C4). The number of stimulus sequences (repetitions) was fixed to 10 times.

3.3. Stimulus-Wise Classification Accuracy

The ROI for the classification was set to C3, Cz, and C4 in the 400–700 ms since the r^2 values indicated obvious changes induced by auditory stimuli. **Figure 6** shows the stimulus-wise classification accuracy for each sound when the auditory stimuli were each presented 10 times. We found that the very negative sound demonstrated the highest accuracy, the very positive sound demonstrated the second highest accuracy, and the positive sound demonstrated the lowest accuracy. **Figure 7** represents the classification accuracy for all 10 stimulus sequences. Two-way repeated-measures ANOVA revealed significant main effects of type of stimulus [$F_{(4,68)} = 2.82, p < 0.05$] and the number of stimulus sequences [$F_{(9,153)} = 48.24, p < 0.001$]. The *post-hoc* pairwise *t*-test revealed that the very negative sound demonstrated the highest accuracy ($p < 0.001$), and the very positive sound showed higher accuracy than the positive sound ($p < 0.01$). In summary, the very negative sound or the very positive sound showed high accuracy.

4. DISCUSSION

To clarify how degrees of valence influence the auditory P300-based BCIs, we applied five affective sounds to an auditory P300-based BCI: very negative, negative, neutral, positive, and very positive. Those sounds had significantly different valence from



each other. ERP analysis revealed that the very negative and very positive sounds showed high r^2 values or the late components of P300. The very negative sound demonstrated significantly higher stimulus-wise classification accuracy than the other sounds. The very positive sound demonstrated the second highest accuracy, which was significantly higher than the positive sound. These results suggest that highly negative or highly positive affective sounds improve the accuracy. As hypothesized, a certain degree of valence is required to influence the BCI.

Our findings show that the very positive sound improved the stimulus-wise classification accuracy, which is in line with our previous study (Onishi et al., 2017). Additionally, we also demonstrated that the very negative sound improved this accuracy. In our previous study, the negative sound and its control demonstrated similar scores on the affective degree of valence, and therefore, no accuracy difference was found. Considering that these two sounds were scored as negative on the valence scale, the results of the current study is consistent with our previous findings.

This study included only one sound for each valence in order to minimize the variance caused by sounds, and to avoid fatigue. This implies that the results may be sound-specific due to the physical properties of sounds. However, it is unlikely that the effect of affective valence was sound-specific given that the stimuli used in this study were different from those used in our previous study (Onishi et al., 2017). Though different physical properties of sounds were evaluated in those two studies, the results obtained were consistent. Moreover, the P300 amplitude

is known to vary with emotional value, which further validates our findings (Johnson, 1988; Kok, 2001).

This study demonstrated that the r^2 values for the late components of P300 displayed by the very positive or very negative sounds were highest around 400–700 ms and centered around the Cz channel. These results were not a simple response to the modality. Wang et al. (2015) examined the r^2 values associated with simple visual gray-white number intensification pronounced number sound, and their combination, and found that the auditory stimuli showed significant r^2 values in fronto-cortical brain regions between 250 and 400 ms, which contradicts our results, while their visual stimuli enhanced the r^2 values mainly around the occipital and parietal brain regions between 300 and 400 ms. Furthermore, they reported that auditory and visual stimuli combined enhanced r^2 values in the large area including parietal area around 300–450 ms. A variety of sounds have been assessed by the point-biserial correlation coefficient. The r^2 values elicited by drums, bass, and keyboard sounds were high within the occipital, parietal, and vertex brain regions between 300 and 600 ms (Treder et al., 2014). Dripping sounds showed Cz channel activity with r^2 values of the P300 (250 to 400 ms) greater than those elicited by beeping sounds (Huang et al., 2018). Japanese vowels, on the other hand, elicited Cz channel activity centered on r^2 values, although different vowels were not presented (Chang et al., 2013). These studies imply that the r^2 values of non-affective auditory BCIs were different from that of affective auditory BCIs. A facial image study demonstrated that upright and inverted facial images resulted in an early peak of r^2 values around 200 ms, although a salient peak was not found for upright facial images (Zhang et al., 2012). A few studies have analyzed affective stimuli using the point-biserial correlation coefficient. One such study analyzed emotional facial images and found large r^2 values at around 400–700 ms, which is also referred to as the late component of P300 or late positive potential (LPP) (Zhao et al., 2011). Affective sounds in our previous study resulted in a late component of P300 at around 300–700 ms; however, this peak was centered around parietal and occipital brain regions (Onishi et al., 2017). These findings suggest that the affective sounds show high r^2 values of the late component of P300. Moreover, the BCI response to affective stimuli may be common amongst modalities given that BCI responses to affective auditory stimuli were similar to those of affective visual stimuli. The effects of multimodal affective stimuli have not been evaluated for use in a BCI, but they would likely elicit ERPs different from those elicited by unimodal stimuli because the brain regions with significant r^2 values were different between different types of stimuli.

As the degree of valence moves away from zero, the r^2 values rapidly increased, implying that the components contributing to classification are not simply in proportion to the valence. A similar tendency was confirmed in previous studies. Steinbeis et al. (2006) evaluated the effect of emotional music on ERPs, focusing on the expectancy violation of chords. The results showed that the amplitude of P300 was enhanced; however, the change was not in proportion to the expectancy. To obtain a late component of P300 with

higher r^2 values, the degree of valence may need to exceed a certain threshold.

The stimulus-wise classification accuracy estimated within the ROI was highest when presenting the very positive and the very negative sounds; however, accuracy was lowest when using the positive sound. These results were unexpected given the amplitudes of late component of P300 and r^2 values. The accuracy may not directly reflect the change of peak amplitude of the component or r^2 values since the stepwise method was applied in addition to target and non-target waveform variance.

Due to experimental constraints, we have estimated the stimulus-wise classification accuracy in response to the ROI using LOOCV. Therefore, we should consider the risk of overfitting because a portion of the information in the test data is used for the spatio-temporal feature selection in LOOCV. The offline stimulus-wise classification accuracy in this study was less than the online classification accuracy (84.1%). Thus, the obvious inflation of the accuracy cannot be confirmed. This tendency is in line with similar previous studies (Guo et al., 2010; Brunner et al., 2011). The overfitting is caused easily when the dimension is high (Blankertz et al., 2011). We think that the current analysis is hard to suffer from overfitting because the simple linear classifier (SWLDA) is further simplified by reducing the spatio-temporal selection using ROI. Moreover, the effect of ROI was controlled in the comparison by equally applying it to all compared conditions.

This study employed all five sounds and evaluated specially designed cross-validation in order to reveal the effect of affective sounds with different degrees of valence. This result can be used for the sound selection for the standard auditory P300-based BCI. When designing multi-command auditory P300-based BCI, very positive or very negative sounds should be employed as much as possible to use the enhanced P300 component, while the neutral sounds should be replaced with the very positive or very negative sounds. However, the mutual effects that occur when employing varieties of sounds in the BCI must be clarified in future studies.

In future studies, methods for classifying auditory BCIs will be necessary. In our previous study, we evaluated the

ensemble convoluted feature extraction method, which took advantage of the averaged ERPs of each sound (Onishi and Nakagawa, 2018). Recently, an algorithm based on tensor decomposition for auditory P300-based BCI was proposed (Zink et al., 2016). This algorithm does not require subject-specific training, which improves the utility of the BCIs. These approaches may help establish a more reliable affective auditory P300-based BCI.

5. CONCLUSION

To clarify how the degrees of valence influence the auditory P300-based BCIs, five sounds with very negative, negative, neutral, positive, and very positive sounds were applied to the P300-based BCI. The very positive and very positive sounds showed higher point-biserial correlation coefficients of late component of P300. In addition, the stimulus-wise classification accuracy of sounds is high for the very positive and very positive sounds. These results imply that the accuracy is not in proportion to the valence; However, it improved when utilizing very positive or very negative sounds.

AUTHOR CONTRIBUTIONS

AO and SN designed the experiment. AO collected the data. AO analyzed the data. AO and SN wrote the manuscript.

FUNDING

This study was supported in part by JSPS KAKENHI grants (16K16477, 18K17667).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2019.00045/full#supplementary-material>

REFERENCES

- Birbaumer, N. and Cohen, L. G. (2007). Brain-computer interfaces: communication and restoration of movement in paralysis. *J. Physiol.* 579, 621–636. doi: 10.1113/jphysiol.2006.125633
- Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048
- Brett, M., Anton, J.-L. L., Valabregue, R., and Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage* 16: 497. doi: 10.1016/S1053-8119(02)90013-3
- Brunner, P., Joshi, S., Briskin, S., Wolpaw, J. R., and Bischof, H. (2011). Does the P300 speller depend on eye gaze? *J. Neural Eng.* 7:056013. doi: 10.1088/1741-2560/7/5/056013
- Chang, M., Makino, S., and Rutkowski, T. M. (2013). “Classification improvement of P300 response based auditory spatial speller brain-computer interface paradigm,” in *IEEE Region 10 Annual International Conference, Proceedings/TENCON* (Xi’an; Shaanxi). doi: 10.1109/TENCON.2013.6718454
- Farwell, L. and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523.
- Finnigan, S., Humphreys, M. S., Dennis, S., and Geffen, G. (2002). ERP ‘old/new’ effects: memory strength and decisional factor(s). *Neuropsychologia* 40, 2288–2304. doi: 10.1016/S0028-3932(02)00113-6
- Furdea, A., Halder, S., Krusienski, D. J., Bross, D., Nijboer, F., Birbaumer, N., and Kübler, A. (2009). An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology* 46, 617–625. doi: 10.1111/j.1469-8986.2008.00783.x
- Guo, J., Gao, S., and Hong, B. (2010). An auditory brain-computer interface using active mental response. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18, 230–235. doi: 10.1109/TNSRE.2010.2047604
- Halder, S., Rea, M., Andreoni, R., Nijboer, F., Hammer, E. M., Kleih, S. C., et al. (2010). An auditory oddball brain-computer interface for binary choices. *Clin. Neurophysiol.* 121, 516–523. doi: 10.1016/j.clinph.2009.11.087
- Höhne, J., Krenzlin, K., Dähne, S., and Tangermann, M. (2012). Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *J. Neural Eng.* 9:045003. doi: 10.1088/1741-2560/9/4/045003

- Höhne, J., Schreuder, M., Blankertz, B., and Tangermann, M. (2010). "Two-dimensional auditory P300 speller with predictive text system," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Buenos Aires), 4185–4188.
- Huang, M., Jin, J., Zhang, Y., Hu, D., and Wang, X. (2018). Usage of drip drops as stimuli in an auditory P300 BCI paradigm. *Cogn. Neurodyn.* 12, 85–94. doi: 10.1007/s11571-017-9456-y
- Ikegami, S., Takano, K., Kondo, K., Saeki, N., and Kansaku, K. (2014). A region-based two-step P300-based brain-computer interface for patients with amyotrophic lateral sclerosis. *Clin. Neurophysiol.* 125, 2305–2312. doi: 10.1016/j.clinph.2014.03.013
- Jin, J., Allison, B. Z., Kaufmann, T., Kübler, A., Zhang, Y., Wang, X., et al. (2012). The changing face of P300 BCIs: a comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement. *PLoS ONE*, 7:e49688. doi: 10.1371/journal.pone.0049688
- Jin, J., Allison, B. Z., Zhang, Y., Wang, X., and Cichocki, A. (2014). An ERP-based BCI using an oddball paradigm with different faces and reduced errors in critical functions. *Int. J. Neural Syst.* 24:1450027. doi: 10.1142/S0129065714500270
- Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: review and synthesis. *Adv. Psychophysiol.* 3, 69–137.
- Kaufmann, T., Schulz, S., Grünzinger, C., and Kübler, A. (2011). Flashing characters with famous faces improves ERP-based brain-computer interface performance. *J. Neural Eng.* 8:056016. doi: 10.1088/1741-2560/8/5/056016
- Klobassa, D. S., Vaughan, T. M., Brunner, P., Schwartz, N. E., Wolpaw, J. R., Neuper, C., et al. (2009). Toward a high-throughput auditory P300-based brain-computer interface. *Clin. Neurophysiol.* 120, 1252–1261. doi: 10.1016/j.clinph.2009.04.019
- Kok, A. (2001). On the utility of P300 amplitude as a measure of processing capacity. *Psychophysiology* 38:557–577. doi: 10.1017/S0048577201990559
- Krusienski, D. J., Sellers, E. W., Bayouh, S., McFarland, D. J., Vaughan, T. M., Wolpaw, J. R., et al. (2006). A comparison of classification techniques for the P300 speller. *J. Neural Eng.* 3, 299–305. doi: 10.1088/1741-2560/3/4/007
- Nagasawa, S. (2002). Improvement of the Scheffé's method for paired comparisons. *Kansei Eng. Int.* 3, 47–56. doi: 10.5057/kei.3.3_47
- Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Onishi, A. and Nakagawa, S. (2018). "Ensemble convoluted feature extraction for affective auditory P300 brain-computer interfaces," in *Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Honolulu, HI).
- Onishi, A., Takano, K., Kawase, T., Ora, H., and Kansaku, K. (2017). Affective stimuli for an auditory P300 brain-computer interface. *Front. Neurosci.* 11:522. doi: 10.3389/fnins.2017.00522
- Pfurtscheller, G., Mueller-Putz, G. R., Scherer, R., and Neuper, C. (2008). Rehabilitation with brain-computer interface systems. *Computer* 41, 58–65. doi: 10.1109/MC.2008.432
- Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Soc. Cogn. Affect. Neurosci.* 2, 67–70. doi: 10.1093/scan/nsm006
- Scheffé, H. (1952). An analysis of variance for paired comparisons. *J. Am. Stat. Assoc.* 47, 381–400.
- Schreuder, M., Blankertz, B., and Tangermann, M. (2010). A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. *PLoS ONE* 5:e9813. doi: 10.1371/journal.pone.0009813
- Sellers, E. W. and Donchin, E. (2006). A P300-based brain-computer interface: initial tests by ALS patients. *Clin. Neurophysiol.* 117, 538–548. doi: 10.1016/j.clinph.2005.06.027
- Simon, N., Käthner, I., Ruf, C. A., Pasqualotto, E., Kübler, A., and Halder, S. (2014). An auditory multiclass brain-computer interface with natural stimuli: usability evaluation with healthy participants and a motor impaired end user. *Front. Hum. Neurosci.* 8:1039. doi: 10.3389/fnhum.2014.01039
- Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380
- Tate, R. F. (1954). Correlation between a discrete and a continuous variable. Point-biserial correlation. *Ann. Math. Stat.* 25, 603–607.
- Townsend, G., LaPallo, B. K., Boulay, C. B., Krusienski, D. J., Frye, G. E., Hauser, C. K., et al. (2010). A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. *Clin. Neurophysiol.* 121, 1109–1120. doi: 10.1016/j.clinph.2010.01.030
- Treder, M. S., Schmidt, N. M., Blankertz, B., Porbadnigk, A. K., Treder, M. S., Blankertz, B., et al. (2014). Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *J. Neural Eng.* 11:026009. doi: 10.1088/1741-2560/11/2/026009
- Wang, F., He, Y., Pan, J., Xie, Q., Yu, R., Zhang, R., et al. (2015). A novel audiovisual brain-computer interface and its application in awareness detection. *Sci. Rep.* 5:9962. doi: 10.1038/srep09962
- Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* 8, 164–173. doi: 10.1109/TRE.2000.847807
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–91. doi: 10.1016/S1388-2457(02)00057-3
- Zhang, Y., Zhao, Q., Jin, J., Wang, X., and Cichocki, A. (2012). A novel BCI based on ERP components sensitive to configural processing of human faces. *J. Neural Eng.* 9:026018. doi: 10.1088/1741-2560/9/2/026018
- Zhao, Q., Onishi, A., Zhang, Y., Cao, J., Zhang, L., and Cichocki, A. (2011). A novel oddball paradigm for affective BCIs using emotional faces as stimuli. *Lecture Notes Comput. Sci.* 7062, 279–286. doi: 10.1007/978-3-642-24955-6_34
- Zink, R., Hunyadi, B., Huffel, S. V., and Vos, M. D. (2016). Tensor-based classification of an auditory mobile BCI without a subject-specific calibration phase. *J. Neural Eng.* 13:026005. doi: 10.1088/1741-2560/13/2/026005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Onishi and Nakagawa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.